MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 2: 1–4
DOI 10.3897/mbmg.2.25649

R package

BBI: an R package for the computation of Benthic Biotic Indices from composition data

Tristan Cordier[1], **Jan Pawlowski**[1]

1 *Department of Genetics and Evolution, University of Geneva, Science III, 4 Boulevard d'Yvoy, 1205 Geneva, Switzerland.*

Corresponding author: *Tristan Cordier* (tristan.cordier@gmail.com)

Academic editor: *Dirk Steinke* | **Received** 2 May 2018 | **Accepted** 25 May 2018 | **Published** 5 July 2018

Abstract

The monitoring of impacts of anthropic activities in marine environments, such as aquaculture, oil-drilling platforms or deep-sea mining, relies on Benthic Biotic Indices (BBI). Several indices have been formalised to reduce the multivariate composition data into a single continuous value that is ascribed to a discrete ecological quality status. Such composition data is traditionally obtained from macrofaunal inventories, which is time-consuming and expertise-demanding. Important efforts are ongoing towards using High-Throughput Sequencing of environmental DNA (eDNA metabarcoding) to replace or complement morpho-taxonomic surveys for routine biomonitoring. The computation of BBI from such composition data is usually being undertaken by practitioners with excel spreadsheets or through custom script. Hence, the updating of reference morpho-taxonomic tables and cross studies comparison could be hampered. Here we introduce the R package BBI for the computation of BBI from composition data, either obtained from traditional morpho-taxonomic inventories or from metabarcoding data. Its aim is to provide an open-source, transparent and centralised method to compute BBI for routine biomonitoring.

Key Words

Biotic indices, biomonitoring, R package, benthos, composition data

Introduction

Biodiversity monitoring is the standard approach for the environmental impact assessment of anthropogenic activities. In marine environments, impact assessments are carried out through benthic macro-invertebrates surveys, which involve the sorting and the morpho-taxonomic identification of numerous specimens for a single site (Borja et al. 2009; Tavakoly et al. 2014). Identified taxa are ascribed to ecological weights that are defined from empirical and experimental data and Benthic Biotic Indices (BBI), such as AMBI (A Marine Biotic Index, Borja et al. 2000), ISI (Indicator Species Index, Rygg 2002), NSI or NQI1 (Norwegian Sensitivity Index and Norwegian Quality Index 1, Rygg and Norling 2013), contain their own set of taxa entries in their database. The relative abundance of each taxa having an ecological weight is being used as input in a formula (usually a weighted sum) to compute the BBI values. This value is continuous but the ecological assessment uses discrete categories (usually in five ordered categories from "very good" to "very bad" quality status), so that the continuous BBI value is often turned into a discrete category in order to make the assessment more "human readable" for regulating agencies and policy-makers. Each BBI contains its own specificities, boundaries between classes and addresses different aspects of biological quality elements. Such disparity prompted the development of a normalised Ecological Quality Ratio (nEQR) within the European Water Framework Directive (WFD), which allows the information provided by these BBI to be concatenated into a single yet integrative index to make impact assessments comparable across countries.

High-throughput amplicon sequencing of environmental DNA (eDNA metabarcoding) offers a fast and cost-effective method to describe biological communi-

Copyright *Tristan Cordier, Jan Pawlowski*. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

ties (Taberlet et al. 2012). Such molecular tools have been successfully used to compute BBI in both freshwaters (Visco et al. 2015) and marine ecosystems (Pawlowski et al. 2014; Chariton et al. 2015; Lejzerowicz et al. 2015). The composition data inferred by the molecular approach is being used in the same way as morphological data to compute BBI, by considering reads abundance as a proxy for species abundance (*sensu* gAMBI, see Aylagas et al. 2014), although the direct comparison is not straightforward, because the abundance of reads is not necessarily reflecting accurately the abundance of the species (Elbrecht and Leese 2015, Vivien et al. 2015, Dowle et al. 2016). Such discrepancies led to the development of correcting factors, using the cell biovolume in the case of diatoms, to lower the effect of such quantification bias (Vasselon et al. 2018) or to the use of machine learning algorithms to bypass the taxonomic assignment step when using metabarcoding data (Cordier et al. 2017).

Some of these BBI can be calculated with user-friendly software (AMBI from AZTI, available at: http://ambi.azti.es), including R packages in the case of freshwaters (see the 'biotic' package, Briers 2016) or marine ecosystems (see the 'BEQI2' package, van Loon et al. 2015). However, none of these packages included the NSI, ISI, NQI1, Bentix or nEQR indices and required the development of custom script in order to use the published reference ecological weights database (Rygg and Norling 2013). Practitioners therefore need to develop their own solutions to make an ecological impact assessment, which could hamper the reproducibility and the cross-study comparison of the results. In addition, as these databases are occasionally updated by their maintainers, all solutions would need to be updated as well, which can further hamper the comparison of the results, both across time and studies. Therefore, an open-source software solution would allow a transparent and centralised method to compute BBI.

Here we introduce the R package BBI for the computation of BBI from composition data, either obtained from traditional morpho-taxonomic inventories or from metabarcoding data. It provides an open-source tool for transparent BBI computation and aims to centralise available tools for BBI in various aquatic ecosystems.

Package description

The R package BBI (version 0.2.0) is available from the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/web/packages/BBI/index.html. It can be installed within the R environment, on any operating system (Linux, macOS or Windows), by using the command install.packages("BBI"). All instructions for installation of current release or development versions can be consulted on the GitHub repository page (https://github.com/trtcrd/BBI). The package requires the package "vegan" (Oksanen et al. 2018).

Table 1. Number of entries per BBI in the reference database.

BBI	Entries	Reference
AMBI	6750	Borja et al. 2000
NSI	585	Rygg and Norling 2013
ITI	1506	Maurer et al. 1999
ISI2012	585	Rygg 2002
Bentix	359	Simboura and Zenetos 2002

A reference dataset is included in the package, containing 7822 metazoan taxonomic entries, covering ecological weights and groups for five BBI (Table 1). The BBI function searches for a match between the taxonomic assignments for the composition data used as input and the morpho-taxonomic reference database (BBI database) and returns a list of objects (Figure 1). These objects include the number of taxa that were found in the database, the values of each BBI per sample and the ecological quality status of samples for each BBI. It also outputs a subset of the original composition data that includes only taxa that had a match in the reference database. The nEQR function of the package can be used to compute the normalised Ecological Quality Ratio (nEQR index) over a set of indices for each sample. An example of usage is available on the GitHub repository page.

Conclusion

We introduced the R package BBI for the computation of Benthic Biotic Indices from composition data. It provides two simple R functions to automate the search for matches between the taxonomic assignments and the reference morpho-taxonomic database of 5 BBI (Table 1), to compute the BBI continuous values for each provided sample, to return the discrete ecological quality status for each pair of sample / BBI and finally to provide the normalised Ecological Quality Ratio index, which normalises the assessment across the BBI.

The BBI package will be kept up-to-date for new entries in morpho-taxonomic reference databases for the 5 BBI included here. Hence, the package aims to provide biomonitoring practitioners with a reliable, up-to-date and open-source tool for the computation of BBI from composition data, either obtained from morpho-taxonomic inventories or by eDNA metabarcoding.

Project description

Title: BBI: an R package for the computation of Benthic Biotic Indices from composition data

Study area description: Microscopy, Metabarcoding, Biomonitoring, Biotic Indices

Download page: https://cran.r-project.org/web/packages/BBI/index.html

Programming language: R

Licence: GNU Affero General Public Licence v3

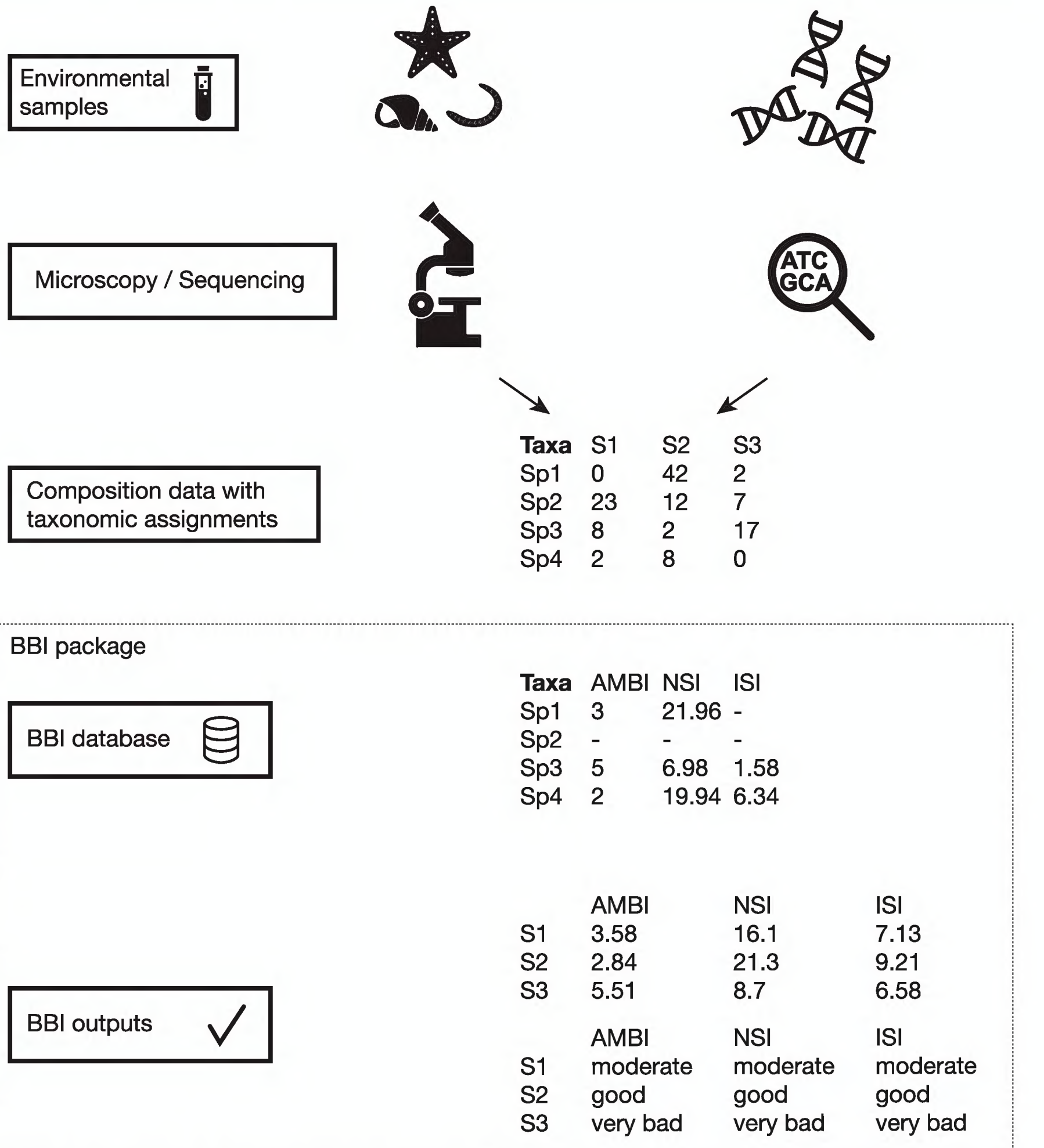

Figure 1. Overview of the BBI package.

Author contributions

Conceived and designed the study: TC; Wrote the R package: TC; Analysed the data: TC; Wrote the paper: TC, JP

Acknowledgements

The R package was developed in the context of the European Cooperation in Science and Technology (COST) Action DNAqua-Net (CA15219). TC and JP were supported by the Norwegian Seafood Research Fund (FHF project number 901092), by the Swiss Network for International Studies and by the Swiss National Science Foundation (grants 316030_150817 and 31003A_159709).

References

Aylagas E, Borja Á, Rodríguez-Ezpeleta N, Borja Á, Rodriguez-Ezpeleta N (2014) Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). PLoS ONE 9(3): e90529. https://doi.org/10.1371/journal.pone.0090529

Borja A, Franco J, Pérez V (2000) A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. Marine Pollution Bulletin 40: 1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8

Borja A, Miles A, Occhipinti-Ambrogi A, Berg T (2009) Current status of macroinvertebrate methods used for assessing the quality of European marine waters: Implementing the Water Framework Directive. Hydrobiologia 633: 181–196. https://doi.org/10.1007/s10750-009-9881-y

Chariton AA, Stephenson S, Morgan MJ, Steven ADL, Colloff MJ, Court LN, Hardy CM (2015) Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environmental Pollution 203: 165–174. https://doi.org/10.1016/j.envpol.2015.03.047

Cordier T, Esling P, Lejzerowicz F, Visco J, Ouadahi A, Martins C, Cedhagen T, Pawlowski J (2017) Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning. Environmental Science & Technology 51(16): 9118–9126. https://doi.org/10.1021/acs.est.7b01518

Dowle EJ, Pochon XC, Banks J, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. Molecular Ecology Resources 16: 1240–1254. https://doi.org/10.1111/1755-0998.12488

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS One 10: 1–16. https://doi.org/10.1371/journal.pone.0130324

Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J (2015) High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Rreports 5: 13932. https://doi.org/10.1038/srep13932

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2018) vegan: Community Ecology Package. R Packag. version 2.4-1. https://doi.org/10.4135/9781412971874.n145

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014) Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecology Resources 14: 1129–1140. https://doi.org/10.1111/1755-0998.12261

Rygg B (2002) Indicator species index for assessing benthic ecological quality in marine waters of Norway. Norwegian Institute for Water Research.

Rygg B, Norling K (2013) Norwegian Sensitivity Index (NSI) for marine macroinvertebrates, and an update of Indicator Species Index (ISI). Norwegian Institute for Water Research.

Simboura N, Zenetos A (2002) Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new Biotic Index. Mediterranean Marine Science 3: 77–111. https://doi.org/10.12681/mms.249

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tavakoly Sany SB, Hashim R, Rezayi M, Salleh A, Safari O (2014) A review of strategies to monitor water and sediment quality for a sustainability assessment of marine environment. Environmental Science and Pollution Research 21: 813–833. https://doi.org/10.1007/s11356-013-2217-5

Vasselon V, Bouchez A, Rimet F, Jacquet S, Trobajo R, Corniquel M, Tapolczai K, Domaizon I (2018) Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution 9(4): 1060–1069. https://doi.org/10.1111/2041-210X.12960

Visco JA, Apothéloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L, Pawlowski J (2015) Environmental monitoring: Inferring the diatom index from next-generation sequencing data. Environmental Science & Technology, 49(13) 49: 7597–7605. https://doi.org/10.1021/es506158m

Vivien R, Wyler S, Lafont M, Pawlowski J (2015) Molecular barcoding of aquatic oligochaetes: Implications for biomonitoring. PLoS One 10(4): e0125485. https://doi.org/10.1371/journal.pone.0125485